Frequent Truth: Impact of Frequency of Misinformation Correction in Extended Extreme Events

Archana Nandakumar
Industrial and Systems Engineering
University of Washington
archanan@uw.edu

Prashanth Rajivan
Industrial and Systems Engineering
University of Washington
prajivan@uw.edu

Abstract

Misinformation management is a growing area of concern in Online Social Network (OSN) organizations. There are several behavioral interventions employed to address misinformation in OSN's. One example is offering users correction when they have engaged with fake news. However, there is little research quantifying the effectiveness of such interventions. We conducted a laboratory experiment to test whether experiencing corrective feedback improved peoples' ability to discriminate true and false news claims during extended extreme events like the COVID-19 pandemic. Participants in the experiment were randomly assigned to four different experiment conditions. Depending on the condition assigned, participants received varying amount of corrective feedback. Results from this experiment suggests that increasing frequency of corrective feedback may not affect peoples' ability to correctly assess information (or misinformation). Political ideology and mistrust in fact-checking organization were found to be the most significant contributing factors. We discuss implications of the findings from this experiment.

1. Introduction

Misinformation is rampant on Online Social Networks (OSNs) spreading either organically through behavioral contagion or on the basis of strategic information operations [1, 2]. Simply put, misinformation is inaccurate information that is believed to be true by a lot of people and is spread on OSNs by the users. Misinformation is not a new phenomenon, but the affordances provided by modern social networks have turbocharged it and have enabled strategic operators to target vulnerable populations [3]. Based on whether a user spreads inaccurate information intentionally or not, becomes the basis for labelling a message as either disinformation (intentional) or misinformation (unintentional) [4]. In this paper, however, the term misinformation is used as an umbrella term to represent all forms of false information.

Misinformation management is a growing area of concern for social media organizations such as Facebook and Twitter. In the event of extended misinformation spread, as is the case during a pandemic such as COVID-19, OSNs employ interventions aimed at tracking and curbing misinformation spread. However, the nature and impact of misinformation seen during the ongoing COVID-19 pandemic has been wide and varied as compared to misinformation during other events such as natural calamities and elections.

Common interventions include using machine learning algorithms to detect and remove misinformation automatically. Early identification of a false news perpetrator or identification of emergence of the new theme of false news aids the organization in managing its spread [5]. Machine learning algorithms and network analysis approaches are usually applied for fact identification, fact extraction, and fact-checking processes. Crowdsourcing for fact-checking is used in conjunction with automation to determine the veracity of a claim made in an article. The partnership of Facebook with Politifact[1] is an example of one such collaboration. Items (articles, social media posts) that these filters automatically flag as rumor or fake are presented to fact-checking experts for verification. Based on their assessment, the item is eventually labeled as fake news or not. The automated filter is then revised based on this expert feedback.

Automation has its limitations, and by the time a message is detected and removed, that message is likely to have been seen and shared by many users on the network [6]. Therefore, there is a growing need to understand the fundamental cognitive challenges people face while discerning truth from the cornucopia of misinformation they experience on social media and develop human-centered interventions. For example, accuracy nudges have demonstrated effectiveness in communicating the risk of misinformation [7].

[1] https://www.politifact.com/facebook-fact-checks/

URI: https://hdl.handle.net/10125/79706
978-0-9981331-5-7

Accuracy Nudges have been shown to encourage users to think about the accuracy of the content as they are reading it. Accuracy nudges was designed to solve the problem of finding a balance between excessive censoring of content and avoiding misinformation.

User education is another broad intervention category that involves training and reinforcing users to use caution when consuming news on social media. This type of intervention focuses on increasing the awareness among users of social networks and internet search engines to develop the practice of verifying sources when reading news on social media. **SIFT** is an example of one such strategy proposed by Mike Caufield [8]. It is an acronym for: **S**top, **I**nvestigate the source, **F**ind better coverage and **T**race claims quotes and media to original context. Each of these steps are tied to some simple skills that can be executed in 30 seconds or less.

Outside the OSN environment, public interest organizations such as local and federal governments, United Nations and associated organizations and other private organizations with the interest of general public also have their interventions to address misinformation. They may issue warnings or fact checks to the public when misinformation on certain topics become exceedingly prevalent and begin to cause societal harm. They may also pass laws punishing perpetrators of misinformation or set standards for Online Social Networks to comply with.

1.1. OSN User Education initiatives

To improve user education and actions, OSNs use warning labels and pre-emptive warnings based on the content of the post. This also includes offering correction to users when they have been found to engage with fake news [9]. Recently, Facebook has started notifying users who have interacted (liked, shared or commented) or have been unknowingly exposed to COVID-19 related misinformation, and redirecting them to factual content about COVID-19 maintained by authoritative sources such as WHO (World Health Organization) [10]. This is a form of corrective feedback provided to users' past actions to engage with misinformation. In fact, different forms of such anti-misinformation feedback are used to warn and correct user engagement with misinformation. For example, OSNs regularly flag and block false messages posted or shared by users, a form of outcome-based feedback. Implementing these interventions that involve the task of identifying misinformation messages that are spreading fast and labeling them with the right warning information relevant to the kind of misinformation that is being spread, involves dedicated resources on part of the OSN organization. Though the effect of providing corrective feedback through warning labels on misinformation detection outcomes has been studied in the past [11], the amount of information and its quality provided in the feedback and the extent to which it makes a difference in the spread of misinformation is unknown. It is necessary to quantify how these interventions that aims to prevent the spread of misinformation perform in extended extreme events such as the COVID-19 pandemic.

1.2. About COVID-19 misinformation

A unique aspect of extended extreme events such as the COVID-19 pandemic is the multi-dimensional nature of misinformation around the subject. The misinformation around COVID-19 is not just restricted to health-related misinformation. Due to the nature of the pandemic, there has been a host of misinformation surrounding this, enough for it to be termed as an 'infodemic'[12]. Action by public authority, spread through the community and general medical claims were among the most common themes of misinformation prevalent in social media.

For the purpose of this research study, the scope was limited to medical misinformation surrounding the COVID-19 pandemic as it has direct health consequences. Surprisingly, even within just this aspect of misinformation, there were several dimensions of misinformation such as misinformation surrounding cures, prevention and the mechanism of spread of the SARS-CoV2 virus.

In this paper, we study the effect of increasing the corrective feedback, similar to that adopted by Facebook [13]. Through this study we would like to understand if increasing corrective feedback to users, will result in better misinformation detection.

2. Background

This section covers the literary background of current research in correction of misinformation and introduces the thesis that will be studied in the paper.

2.1. Misinformation Correction

Several meta-analysis studies have been conducted aggregating results from multiple papers that conducted behavioral experiments on correction of misinformation. They give us a general direction on aspects of correction that have worked effectively to curb the spread of misinformation in the past [14, 15, 16]. Actively

correcting users after the act using corrective rebuttals in different formats or dissuading them before they engage in misinformation with warnings have typically been the focus of corrective studies in misinformation. The study by Walter and Murphy (2018), suggests that de-biasing technique, platform design and delay of corrections have a statistically significant effect on the spread of misinformation, suggesting that better platform designs coupled with the right fact-checked correction in a timely manner has the capability to slow the spread of misinformation. This particular meta-analysis looked at misinformation from across domains, so the findings are not restricted to any one kind of misinformation. Where timeliness of correction is concerned, the meta-analyses suggest that while delays that are small (with time enough to complete a filler task before given a correction), the difference in impact is not that great but if the corrections are delayed by more than a day, the results are significantly different. This suggests that timeliness of correction is important though immediacy may not be that important. Warnings have been suggested as an intervention in several studies (at least 3 studies before 2018 as per the meta-analysis by Walter and Murphy 2018). They have been shown to be effective when compared to claims on social media without warnings. In fact, one study [17], conducted a simulation using their experiment results, suggested that claims without warnings spread to 5 times more people than claims with warnings.

The meta-analyses studies suggest that more systematic rebuttals to claims are more effective than mere warnings that appear before claims. A detailed debunking of a claim enables the user to abandon their beliefs on the rumor by providing an alternate coherent model of belief[14]. In fact, detailed fact-checking and coherence related de-biasing methods worked better in several studies than just pointing out the credibility of the source. The observed difference in impact in these methods were statistically significant[15].

Thus, existing studies suggest the use of timely correction of misinformation using a correctly worded rebuttal, would be effective in slowing the spread of misinformation and in general educating the public about the pandemic. Considering the insights on timeliness, it would suggest an immediate algorithmic correction from OSNs on misinformation from a crowdsourced fact-check would be effective.

2.2. Frequency of corrections

Though the effect of providing corrective feedback on misinformation detection outcomes has been studied in the past [11], the amount of information and its quality provided in the feedback is unknown. As the authors in Byrd and John (2021) noted, combining feedback training with a critical thinking intervention is necessary for the users to learn the right behavior. Getting the users to "learn" the right behavior is an important aspect of misinformation management. According to the instance-based learning theory (IBLT) [18], frequent experience of events will reinforce the key ideas in memory and enable people to make the correct judgments. In this paper, we investigate whether the frequency of experiencing misinformation corrections and information reinforcements influences the more complex detection decisions in social media news feeds. If lessons from detection problems in Cybersecurity were to be learned, it can be expected that when the frequency of corrective feedback increases, there will be an increase in people's attention that would result in higher sensitivity to misinformation [19]. A test of this hypothesis would provide the theoretical grounds of the frequency with which OSN users could be provided with algorithmic corrections or feedback in the OSN platform to address the problem misinformation from the perspective of user behavior.

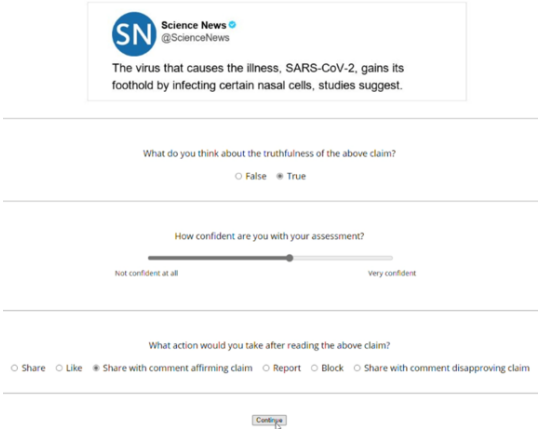

Figure 1. Example trial in the experiment presenting a news claim.

3. Experiment Design

Using an experimental design consisting of three phases, we measured whether experiencing frequent corrective feedback will enable people to learn to become wary of engaging with misinformation. The experiment had 3 phases: a pre-phase, an intervention phase and a post-phase. The pre-phase and post-phase had 15 trials each. The intervention phase had 24 trials. The number of trials per phase were determined through several pilot tests. The aim was to limit the

experiment duration to a reasonable limit, about 30 minutes. During each trial, a news claim designed to look like a tweet was presented to the human participant (see Figure 1). Below each claim, the participants were asked to provide their assessment of the claim, whether what was mentioned in the claim was true or false, how confident they were with their assessment and the action they would take after reading the claim (like, share, mute etc.). During the intervention phase of the experiment, some of participants received corrective feedback to some of their decisions. Depending on the experiment condition assigned, participants either received no feedback or increasing levels of corrective feedback. The corrective feedback was presented as a correction from the fact-checking organization Politifact. Participants did not receive any corrections during the pre- and post-phases of the experiment. Such an experiment design was used to study the following aspects about misinformation correction.

3.1. Discerning information

Does offering more corrective feedback result in improved discerning of information? The experiment consisted of four conditions. In the control condition, the participant does not receive any corrective feedback. However, participants in the other three conditions received progressively higher amounts of corrective feedback (25%, 50% and 75%) during the intervention phase. They received corrective feedback on both information and misinformation claims. Participants' decisions during the pre- and post phase were compared to measure the impact of frequency of corrective feedback. The hypothesis was that increasing the frequency of corrective feedback will improve the performance of the participant in being able to discern and differentiate between misinformation and information in claims. For example, to counter misinformation spread regarding effectiveness of masks, does it help if organizations such as Facebook provided frequent corrective feedback to the users who engage with misinformation regarding the use of masks to prevent COVID.

3.2. Assessments and actions

Do assessments correctly translate into actions?

Earlier studies of motivation in users sharing misinformation found that a considerable portion of users shared misinformation for reasons other than their opinion of its truthfulness [20]. Some of the reasons for sharing misinformation were to obtain others' opinions on that information, to express own opinions, and to interact with others. These results suggest that there may be some mismatch between what the user thinks of the correctness of the news claim versus how the user chooses to amplify the claim. For this reason the user's assessment of the claim and the action chosen for the claim was analyzed to study gap between belief and behavior, for example, when a user retweets unproven prevention methods for COVID, do they really believe in those methods?

3.3. Correction performance by topic

Did the users perform better in some dimensions of COVID-19 misinformation/information?

The experiment is designed such that the participants were provided corrective feedback on randomly chosen information and misinformation claims (without controlling for the dimension of COVID medical misinformation for which they received feedback). The experiment design was used to measure if participants' assessments of claims for certain dimensions were better than others. For example, were participants better at identifying fake cures than they were at identifying claims on COVID-19 disease spread.

3.4. Behavior learning

Does offering more corrective feedback result in users learning the right behavior?

Offering a well-worded rebuttal for misinformation is cited as one of the most effective methods for combating misinformation in the meta-analyses. Thus, the rebuttals were worded in the form of a Politifact Fact-check with their trademark Truth-o-meter[2]. In the cases of misinformation, the feedback was also carefully worded in the form of a "truth sandwich" [21] by starting with the truth, indicating the lie in the claim and returning and reiterating the truth. In this way, it was passively suggested to the users who received corrective feedback that a fact-checking service such as Politifact may be used to verify the information on a given claim. The participants' responses to the claims in pre-phase and post-phase were compared to measure if there were any learning effects from the intervention phase. For example, when offered more corrective feedback do users imbibe the behavior of researching the facts of a claim when in doubt.

4. Methods

The experiment interface was developed using JsPsych, which is a javascript library for developing behavioral experiments in a web browser [22]. Amazon

[2]https://www.politifact.com/truth-o-meter/media/

MTurk was used to recruit participants for the experiment. The main constraints for being a participant in the study was being over 18 years old, geographical location (constrained to the United States) and the requirement that they should have a Twitter account (so that they would be familiar with a typical social media environment). In addition to this, it was also required that the participants not be associated with the University of Washington. All experiment protocols were approved by the university institutional review board (IRB).

4.1. Participant Characteristics

200 participants recruited from Amazon Mechanical Turk were randomly assigned to one of the four experimental conditions (No feedback, 25%, 50% and 75%). Care was taken to ensure we received responses from equal number of participants (50) in each condition. 17 out of the 200 participants failed attention checks and their data was excluded from the study. Of the remaining 183 participants, 84 (42.2%) were female and 115 (57.7%) were male. 43 (21.6%) were between the ages 18-29 and 156 (78.4%) were between the ages 30-49. Their political ideologies were collected on a scale of 1 to 7 (1 being most liberal and 7 being most conservative). About 117 (58.7%) had left wing ideologies, 57 (28.6%) had right wing ideologies and the rest were neutral.

4.2. Claims Used

News claims that were displayed to the participants were collected from real instances of information and misinformation. True claims from various news categories were collected: Politics, Business, Sports, Science and Technology and COVID-19. The goal was to present the participant with news claims from different categories to simulate a real-world news feed in social media. However, all the misinformation claims in all three phases of the experimented were specific to COVID-19 topics, for e.g., misinformation claims regarding Covid-19 spread and cures. An experiment design with only Covid-19 related claims (true and misinformation) would bias participants into producing more false alerts - classifying majority of claims as misinformation. Furthermore, such a design is critical for measuring how participants discriminated true news from misinformation using signal detection theory [23] - to measure how well participants discriminated signals (Covid-19 misinformation claims) from noise (true claims sampled from different news categories including Covid-19 related news).

For COVID-19 related misinformation and information tweets, the International Fact Checking Network's (IFCN) aggregator website for COVID-19 fact-checks was used[3]. For other news categories, the Associated Press's (AP)[4] twitter handle for every news category (for example, @ap_politics for politics news) was mined to gather salient information in that category. Misinformation for these categories were identified by searching Snopes.com and Politifact. Real instances of tweets amplifying a particular information were collected using the Hoaxy service [24, 25]. With this, a database of about 100 news claims was built. This was reduced to 54 claims and 4 additional claims were included as attention checks. Therefore, a total of 58 claims were used in the experiment to keep experiment duration within reasonable limits. The Python Imaging Library was used to design these claims to look like real tweets from social media. The authors of these news claims were made to appear as a mixture of verified and unverified personalities on social media. In all three phases, one-third of all the messages were misinformation.

4.3. Procedure

When the participants click on the survey link in MTurk, they were redirected to consent page where they entered their MTurk id, age and declared non-association with University of Washington. Then they were redirected to the instructions page of the experiments which gave them details about the three phases and the tasks expected from them. Participants in the control experience one long phase with no corrective feedback. Participants in all other conditions completed the experiment in phases. They first made judgments on the claims presented in pre-phase and then, were moved to intervention phase, where depending on their randomly assigned experimental condition, they received action-based feedback on a certain percentage of randomly selected news claims. Example of a corrective feedback is given in Figure 2. In the figure the participant choose to amplify a misinformation claim and is presented with a correction message from Politifact. The feedback text was worded as a "truth sandwich". After the intervention phase, the participants were moved to the third and final post-phase. For every claim shown to the participant, in addition to their responses on the claim (their assessments, confidence level and actions), we also measured their response time, recorded whether feedback was provided to a particular claim in the intervention phase and whether the action they chose was appropriate. The appropriate

[3] https://www.poynter.org/ifcn-covid-19-misinformation/
[4] https://twitter.com/AP

actions would be to choose options that do not spread misinformation and to choose options that do not curb information). At the end of post-phase, the participants were presented with a final survey that collected their demographic data which included questions on age, gender, education, political ideology, concern about COVID, how often they check COVID related news, the different social media websites they are familiar with and how much they trust Politifact. All participants were paid $4 as a fixed base payment and received a $1 bonus payment for passing all attention checks.

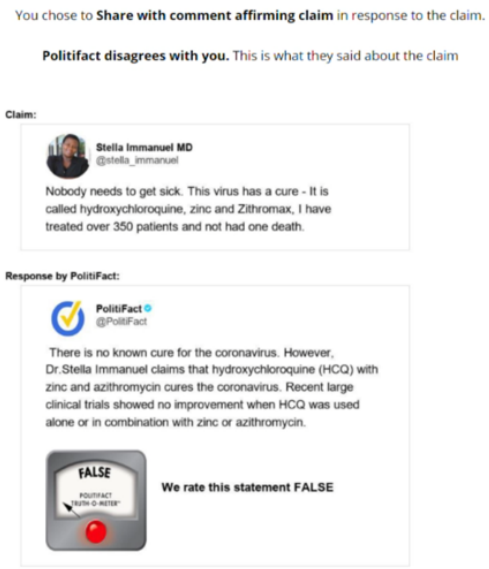

Figure 2. An example of negative feedback given to the participant

5. Results

The design of this experiment can be best described as a mixed-effects model. Mixed models are a good fit here because we collected multiple measurements from each participant and mixed models are well suited for such within-subject designs. In this study, we are interested in the effect frequency of feedback which we tested as 4 different experimental conditions (0%, 25%, 50% and 75% feedback conditions) with about 50 users in each condition. Hence, for our study feedback frequency is modelled as a Fixed effect. Random effects in our study are demographic variables and other variables that capture individual differences.

5.1. Effect of corrective feedback

In this experiment, we are interested in how well a participant discerns information (and misinformation). For every news claim shown, we gathered data on whether the user has assessed a piece of information correctly and whether they then take the right action towards curbing the spread of misinformation and amplifying the right information. The two outcomes were named 'correct assessment' and 'correct action'. Figure 3 shows a histogram of the 'correct action' outcome by experiment condition and phase. Figure 4 shows a histogram of the 'correct assessment' outcome by experiment condition and phase. We found a significant correlation between participants' actions and their assessment of the information (over 88%) which suggested that the remaining analysis could focus on one of these two measures.

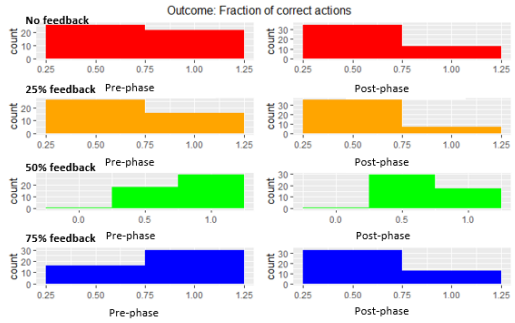

Figure 3. A histogram of fraction of correct actions by experiment phase and condition

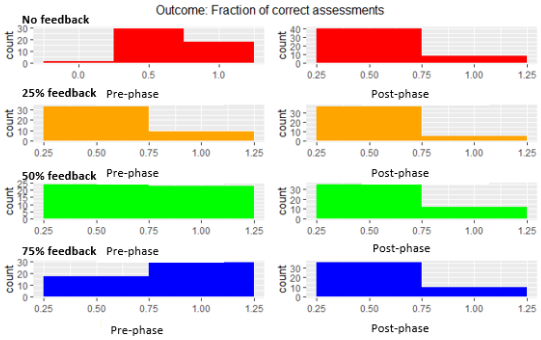

Figure 4. An histogram of correct assessments by experiment phase and condition

From an initial examination of the outcomes, it

seems that corrective feedback has a mixed effect. For further analysis, a mixed model of the participants' percentage of correct assessments from the experiment was computed with feedback condition as the primary fixed effect and news category and their interaction with feedback condition as secondary fixed effects and subject-wise variability as a random effect. The results from this analysis are tabulated in Table 1.

Table 1. Results from Mixed Model Analysis

Effects	Estimate	t-value	p-value
Intercept	0.63	23.43	0.001
Feedback %	≤ 0.001	1.37	0.17
Celeb news	-0.34	-8.48	0.001
COVID Cures	0.14	4.32	0.001
COVID Prevention	0.15	5.23	0.001
COVID Spread	-0.057	-1.91	0.06
Politics news	-0.118	-3.57	0.001
Science and tech	0.167	4.14	0.001

From the table, it can be seen that the feedback condition is not a statistically significant factor in determining the correct assessment percentage. However, the different categories of news affect the outcomes. Particularly, participants were more likely to make errors in discerning celebrity news, news claims containing information / misinformation about the mechanism of spread of COVID (though this effect was not statistically significant). Participants were less likely to make errors when discerning news claims containing science or technology news and news about the prevention of COVID.

With the model results suggesting that increasing corrective feedback does not help the user discern information/misinformation better, an important consideration is whether time elapsed was an important factor in the performance of the user. It is possible that exhaustion associated with reading and evaluating multiple news claims may have lead to a decreased performance in the post-phase.

5.2. Learning effects

For the primary outcome of correct assessment, we calculated the signal detection outcomes d-prime (a measure of sensitivity) and response bias (bias toward classifying a claim as misinformation or information) to study how the experiment affected the participants' discerning abilities. The average d-prime by phase and experiment condition is provided in Table 2. The average response bias by phase and experiment is provided in Table 3.

With d-prime for assessments, apart from the anomaly for the 25% feedback condition, in all other cases the d-prime decreases in the post-phase condition as compared to the pre-phase condition. This implies that the difference between the hit rate and the false alarm rate reduces and that the user discerns information better in the pre-phase than the post-phase. Similarly, with response bias or beta, we find that for both actions and assessments, the bias to say 'yes' or classify a news as misinformation increases between pre and post phases irrespective of the type of feedback condition. A t-test conducted pre-phase and post-phase signal detection measures for the different experiment conditions yielded a statistically significant difference only for the 75% feedback condition. From the trend, we can draw the implication that as the participant receives more corrective feedback on their actions, they become more conservative in their assessments, erring on the side of caution.

Table 2. Average d-prime

Condition	Pre-phase	Post-phase
No feedback (0%)	1.20	0.95
25%	0.80	1.00
50%	1.20	1.04
75%	1.55	1.16

Table 3. Average response bias

Condition	Pre-phase	Post-phase
No feedback (0%)	0.86	0.73
25%	0.80	0.78
50%	0.89	0.72
75%	0.79	0.64

5.3. Demographic effects

The demographic effects were studied by using a generalized linear model of demographic factors collected from participants to predict the signal detection outcomes in the post-phase. The factors considered in the analysis were experiment condition, age, gender, political ideology (scale of 1 to 7, 1 being very liberal and 7 being very conservative) and mistrust in the credibility of Politifact (0 - Very trustworthy to 4 - Not trustworthy at all). Table 4 shows the estimates of the model and their statistical significance.

From Table 4, it can be seen that the most significant demographic factors were Political ideology and Mistrust in the fact-checking organization employed in the study. Participants who subscribed to right-wing political ideology were likely to have a lower sensitivity to discerning information (or misinformation). Also participants who reported higher mistrust in Politifact

Table 4. Generalized Linear Model predicting post-phase d-prime values by demographic factors

Effects	Estimate	t-value	p-val
Intercept	1.582	7.66	0.001
Feedback %	≤ 0.001	-0.22	0.82
Gender: Male	0.21	1.99	0.05
Age	-0.07	-0.52	0.60
Political ideology (P)	-0.095	-2.07	0.04
MistrustinPolitifact (MP)	-0.21	-1.89	0.05
Interaction (P * MP)	-0.02	-0.69	0.49

and their credibility were likely to demonstrate lower sensitivity. It can also be noted that Political ideologies and mistrust in Politifact did not have any interaction effects. This implies that mistrust in the fact-checking organization was prevalent across the spectrum of Political ideologies.

6. Discussion

6.1. Feedback and nuance

Results from the experiment suggest that frequent corrective feedback may not have a significant effect on users' behavior. The expectation was that increased frequency of corrective feedback will emphasize the importance of fact-checking when in doubt and will become a behavior adopted by the participants. But most participants in the experiment made judgments on the claims presented to them in a vacuum. This is evidenced by how the users responded to the question "How often did you look up on the internet for the validity of the claims shown in the experiment?". 172 participants out of 183 (94%) reported not checking the facts. This issue, particularly with MTurk participants, has been discussed previously [26].

Eliciting response from the participant for every news claim is a key difference between the experimental environment and the social media environment. In a typical OSN environment, users can sift through plenty of information without interacting with anything. But in the experimental environment, they were required to give their assessment on every single claim that they read. This may have caused time constraints which might have prevented the participants from looking up the veracity of claims.

One possible explanation for the lack of difference is information avoidance [27]. Expectation is that people when presented with useful information would take that information into consideration while making future judgments. However, decades of research has shown that people dislike receiving information that conflicts with their existing beliefs and mental models [28]. This could lead to continued influence effect of misinformation on individual's decisions [29]. The other possible explanation is that people may have difficulty transferring correction received on one misinformation claim to correcting another, albeit thematically related, misinformation. It is also important to note that participants in the experiment did better on specific news categories as compared to others. Particularly, claims on COVID prevention and cures and Science and technology news were categories on which participants did well. Politics and Celebrity news were topics where participants did not do well. These differences arise due to the nature of news in these categories and has been noted in previous studies as well [15]. Future research must analyze the impact of frequent correction on specific claims.

6.2. Learning

Results from the experiment reveal that exposing users to more corrective feedback may decrease their sensitivity and they may become more biased towards classifying a claim as misinformation. This is typically how detection task outcomes are affected [30, 31]. The approach of increasing feedback works for most detection problems (such as baggage screening and malicious email detection), where a slight increase in false alarms does not adversely affect outcomes. However in information dissemination, identification of misinformation is as important as not discrediting facts. In reviewing survey data, it was seen that most participants (94%), regardless of frequency of corrective feedback, did not look up the facts when presented with a claim. This implies that simply exposing users to more frequent misinformation feedback, even if it is a correctly worded rebuttal, is creating experiences without adding contextual meaning to those experiences. The feedback given must be followed up with more understanding of how the process of information discerning works, so that users imbibe the right behaviour when consuming news on social media.

6.3. Limitations and Future work

One of the limitations of the study in terms of the participant demographics was the limited age range and political ideologies of the participants. 78.4% of the participants were between the ages 30 and 49, and only 28.6% of the participants had a right-wing political ideology. The lower percentage of participants with right wing political ideology may also explain the lack of interaction between political ideology and mistrust in Politifact despite their statistically significant effects on

the main outcome. Future work in this area has to focus on methods to include participants across different age ranges and political ideologies.

Despite having restricted the theme of information to COVID-19 disease related claims, the authors found that claims within this domain itself was wide and varied. There were claims pertaining to mechanism of spread of the disease, claims pertaining to disease prevention and those discussing cures and treatments. Due to the variation in performance of participants in each type of claim, it was not possible to gain insights on the effect of feedback frequency on any one type of claim. Future studies can focus on one aspect of COVID-19 medical misinformation. Alternatively, other dimensions of COVID-19 misinformation like Xenophobic or Racial prejudices can be explored [32].

Another limitation as discussed previously was the requirement to collect data from participants after every claim which might have exhausted the participants. In future work, the experimental environment can incorporate some OSN environment features such as browsing without having to make assessments and choose actions for every claim, so that the experimental environment closely mirrors the real life experience. The insight from demographic factors, particularly that of mistrust in the fact-checking organization denotes the larger problem of information avoidance when the information presented in not in line with the beliefs of the participant. This has been observed in previous work, where in users do not change their beliefs even when repeatedly presented with facts [33]. In the future, these differences can be addressed by gathering their ideologies beforehand and having influencers or organizations that embrace their political beliefs reiterate facts especially where medical misinformation is concerned.

References

[1] C. T. Bergstrom and J. B. Bak-Coleman, "Information gerrymandering in social networks skews collective decision-making," *Nature*, vol. 573, pp. 40–41, sep 2019.

[2] A. J. Stewart, M. Mosleh, M. Diakonova, A. A. Arechar, D. G. Rand, and J. B. Plotkin, "Information gerrymandering and undemocratic decisions," *Nature*, vol. 573, pp. 117–121, sep 2019.

[3] K. Starbird, A. Arif, and T. Wilson, "Disinformation as collaborative work," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, pp. 1–26, nov 2019.

[4] V. F. Hendricks and M. Vestergaard, *Reality Lost*. Springer International Publishing, 2019.

[5] M. Fernandez and H. Alani, "Online misinformation: Challenges and future directions," in *Companion Proceedings of the The Web Conference 2018*, WWW '18, (Republic and Canton of Geneva, CHE), p. 595–602, International World Wide Web Conferences Steering Committee, 2018.

[6] P. M. Krafft and E. S. Spiro, "Keeping rumors in proportion," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, ACM Press, 2019.

[7] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand, "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention," mar 2020.

[8] M. Caufield, "Sift [stop, investigate the source, find trusted coverage, and trace claims]," June 2019 [Online].

[9] N. Mele, D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs, and C. Mattsson, "Combating fake news: An agenda for research and action," 2017.

[10] R. Guy, "An update on our work to keep people informed and limit misinformation about covid-19." Facebook Blog, Apr. 26, 2020 [Online].

[11] K. Byrd and R. John, "Tell me the truth: Separating fact from fiction in social media following extreme events," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, p. 2718, 2021.

[12] J. S. Brennen, F. Simon, P. N. Howard, and R. K. Nielsen, "Types, sources, and claims of covid-19 misinformation," *Reuters Institute*, vol. 7, pp. 3–1, 2020.

[13] B. Perrigo, "Facebook is notifying users who have shared coronavirus misinformation. could it do the same for politics?." Time.com, April 16 2020 [Online].

[14] M.-p. S. Chan, C. R. Jones, K. Hall Jamieson, and D. Albarracín, "Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation," *Psychological science*, vol. 28, no. 11, pp. 1531–1546, 2017.

[15] N. Walter and S. T. Murphy, "How to unring the bell: A meta-analytic approach to correction of misinformation," *Communication Monographs*, vol. 85, no. 3, pp. 423–441, 2018.

[16] N. Walter, J. J. Brooks, C. J. Saucier, and S. Suresh, "Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis," *Health Communication*, pp. 1–9, 2020.

[17] P. Ozturk, H. Li, and Y. Sakamoto, "Combating rumor spread on social media: The effectiveness of refutation and warning," in *2015 48th Hawaii International Conference on System Sciences*, pp. 2406–2414, IEEE, 2015.

[18] C. Gonzalez, J. F. Lerch, and C. Lebiere, "Instance-based learning in dynamic decision making," *Cognitive Science*, vol. 27, no. 4, pp. 591–635, 2003.

[19] B. D. Sawyer and P. A. Hancock, "Hacking the human: the prevalence paradox in cybersecurity," *Human factors*, vol. 60, no. 5, pp. 597–609, 2018.

[20] X. Chen, S.-C. J. Sin, Y.-L. Theng, and C. S. Lee, "Why do social media users share misinformation?," in *Proceedings of the 15th ACM/IEEE-CS joint conference on digital libraries*, pp. 111–114, 2015.

[21] R. Clark, "How to serve up a tasty 'truth sandwich?'." poynter.org, August 18 2020 [Online].

[22] J. R. De Leeuw, "jspsych: A javascript library for creating behavioral experiments in a web browser," *Behavior research methods*, vol. 47, no. 1, pp. 1–12, 2015.

[23] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behavior research methods, instruments, & computers*, vol. 31, no. 1, pp. 137–149, 1999.

[24] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer, "Hoaxy: A platform for tracking online misinformation," in *Proceedings of the 25th international conference companion on world wide web*, pp. 745–750, 2016.

[25] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nature communications*, vol. 9, no. 1, pp. 1–9, 2018.

[26] E. A. Necka, S. Cacioppo, G. J. Norman, and J. T. Cacioppo, "Measuring the prevalence of problematic respondent behaviors among mturk, campus, and community participants," *PloS one*, vol. 11, no. 6, p. e0157732, 2016.

[27] R. Golman, D. Hagmann, and G. Loewenstein, "Information avoidance," *Journal of Economic Literature*, vol. 55, no. 1, pp. 96–135, 2017.

[28] R. P. Abelson, E. E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, and P. H. Tannenbaum, "Theories of cognitive consistency: A sourcebook.," 1968.

[29] S. Lewandowsky, U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook, "Misinformation and its correction: Continued influence and successful debiasing," *Psychological science in the public interest*, vol. 13, no. 3, pp. 106–131, 2012.

[30] K. Singh, P. Aggarwal, P. Rajivan, and C. Gonzalez, "Training to detect phishing emails: Effects of the frequency of experienced phishing emails," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, pp. 453–457, SAGE Publications Sage CA: Los Angeles, CA, 2019.

[31] P. Madhavan, C. Gonzalez, and F. C. Lacson, "Differential base rate training influences detection of novel targets in a complex visual inspection task," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 51, pp. 392–396, SAGE Publications Sage CA: Los Angeles, CA, 2007.

[32] H. Budhwani and R. Sun, "Creating COVID-19 stigma by referencing the novel coronavirus as the "chinese virus" on twitter: Quantitative analysis of social media data," *Journal of Medical Internet Research*, vol. 22, p. e19301, may 2020.

[33] E. Thorson, "Belief echoes: The persistent effects of corrected misinformation," *Political Communication*, vol. 33, no. 3, pp. 460–480, 2016.